Acta Cryst. (1993). D**49**, 158–167

Ab Initio Direct Methods: Practical Advice for Getting Beyond the First 300 Atoms

BY DAVID A. LANGS

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(*Received* 27 *April* 1992; *accepted* 9 *June* 1992)

Abstract

Not all crystallographic structural investigations are amenable to a phasing solution by direct methods alone. Guideline procedures are outlined which are intended to help the evaluation of whether direct-methods procedures may be expected to phase diffraction data for large molecular structures. This analysis is directed at three separate levels of inquiry: (1) How good are the primary data and can E values be derived to represent a point-atom structure. (2) How well do the data interact through phase relationships and may they be expected to produce a stable phasing solution. (3) What is the prognosis for finding recognizable solutions. Data are presented from the post-mortem analyses of a number of large, difficult-to-solve, structures to illustrate each of these points. Direct-methods practitioners are to be encouraged that crystal structures having more than 300 atoms per asymmetric unit may occasionally be determined utilizing present methodologies provided that an *a priori* prognosis for obtaining a solution is favorably high, adequate computational resources are available, and sufficient persistent effort is applied.

Introduction

More than 40 years have passed since the first successful applications of direct phasing methods to crystal structure determination were reported in the earliest volumes of this journal (Gillis, 1948; Kasper, Lucht & Harker, 1949). Most crystallographers, however, were not convinced that these methods contained any new phasing information which could not have been obtained from an appropriate interpretation of the Patterson function. The claim that the centrosymmetric phase problem was, in principle, solved (Hauptman & Karle, 1953, 1954), was received with guarded skepticism (Cochran & Woolfson, 1954; Vand & Pepinsky, 1954; Rogers & Wilson, 1955). It was probably not until more complex centrosymmetric structures were determined (Karle, Hauptman, Karle & Wing, 1958) that most skeptics were willing to conclude that although these techniques could not absolutely guarantee a solution to the phase problem, they could solve structures that were indeterminable by any other method. Subsequent applications to non-centrosymmetric structures further underscored this truth (Karle & Karle, 1964, 1966; Karle, Karle & Estlin, 1967; Oh & Maslen, 1968) and triggered urgent demands for computer software from researchers who wished to perform these state-of-the-art structural analyses on an enormous class of chemical substances and natural products which could not have been determined by any other known method.

It took nearly 20 years from the time that Harker & Kasper first reported the inequality relationships (Harker & Karper, 1948) to develop practical phasing strategies and the computational technology which permitted researchers to exploit direct methods in a routine, black-box, manner. These direct-methods phasing strategies and computing algorithms have improved over the intervening 25 years where it is now fairly routine to expect the majority of non-centrosymmetric structures having up to 100 non-hydrogen light atoms in the asymmetric unit to be phased without too much difficulty, provided one has a sufficient number of accurately measured diffraction data. Routine applications to structures possessing more than 100 atoms remain, for the time being, less certain, and a large fraction of these structures may require a number of solution attempts to reach a successful conclusion.

Our present experience does not permit us to reliably extrapolate the limit of structural complexity which may ultimately be achieved by using these methods (see for example, Cochran, 1958). Only a small number of the unknown light-atom structures possessing more than 200 atoms have been successfully determined using traditional tangent-formula procedures (Karle, Karle, Mastropaolo, Camerman & Camerman, 1983; Langs, 1988). The Sayre tangent formula (Woolfson & Yao, 1990) and minimal function (Guo & Hauptman, 1990), discussed in this symposium, appear to have a larger radius of convergence than traditional tangent-formula methods, and it is expected that more structures exceeding 200 atoms will be determined in the near future. Other tangent-formula modifications such as those incorporating higher-order determinants (Karle, 1971; Tsoucaris, 1980) may eventually prove to be more powerful in *ab initio* macromolecular phasing

0907-4449/93/010158-10$06.00

applications, but, in the words of Jerome Karle (1989), 'These techniques have not achieved widespread use up to this time, presumably because there are alternative techniques that investigators have found to be preferable.'

Entropy-based phasing methods, in comparison, appear to be capable of converging to a solution for much larger structures, say 600 atoms or more (Harrison, 1989; Sjölin, Prince, Svensson & Gilliland, 1991; Sjölin & Prince, 1991; Gilmore, Henderson & Bricogne, 1991), but the success of these procedures has not always been well documented and may require sieving through a large number of trial electron-density maps in order to find one which is capable of producing a solution. The Sayre tangent formula and minimal functions may be useful for producing these trial maps in application to small protein structures. A factorialization scheme to sample the phase space of zonal restricted data in order to produce these trial maps for larger macromolecules has been demonstrated for a known protein structure containing 2800 non-hydrogen atoms (Sjölin, Prince, Svensson & Gilliland, 1991), but discussions in the course of this symposium have questioned whether this achievement could be successfully repeated for some unknown macromolecular structure at this time.

This paper will provide a review of a number of important criteria which are vital to the success of phase-invariant-based direct-methods procedures. It will be shown that even the best direct-methods strategy can be defeated if one is blind to certain obvious limitations that may prevent a phase solution from being obtained. Often these difficulties can be assessed beforehand, and may influence one to alter the phasing strategy, attempt to measure better data, or abandon the problem altogether.

The primary data

Small-molecule direct methods have conditioned crystallographers to focus primarily on two sources of potential difficulty concerning the primary intensity data, but it will be necessary to address these and other less publicized concerns when considering phasing applications to larger structures. Firstly, how well can a set of normalized E values be scaled, and secondly, how might measuring errors in the X-ray intensities affect these data, particularly the largest and smallest E magnitudes. The phases of the largest E values are actively used in the tangent formula; the smallest E's are normally used to establish a set of Ψ_0 triples or negative quartet relationships which are passively used to compute figures of merit to help indicate correct phase solutions. In this latter regard one is cautioned to measure even the weakest data with reasonable precision.

Table 1. *E-statistical analysis for the $P2_12_12_1$ structure of hexadecaisoleucinomycin ($N = 127$ atoms)*

A total of 6429 data were measured, scaled and sorted into ten approximately equal population shells out to a maximum $\sin\theta/\lambda$ of 0.50 Å^{-1}. Column 2 gives the number of reflections in each shell, column 3 the fraction of the data for which $F > 3\sigma(F)$, column 4 the fraction of $|E| < 0.70$, column 5 the fraction of $|E| > 1.27$, and column 6 gives the average value of $|E|^2$ for the shell. The fraction of $|E| < 0.70$ in the tenth or highest $\sin\theta/\lambda$ shell dips markedly from 0.44 to 0.35, largely due to the decreasing percentage of data observed at the $3\sigma(F)$ level, *i.e.* from 0.299 to 0.206. This indicates that the accuracy of the largest and smallest $|E|$ amplitudes has degraded.

Shell No.	No. of E's	$> 3\sigma$	< 0.7	> 1.27	$\langle E^2 \rangle$
1	720	0.968	0.456	0.193	0.998
2	669	0.936	0.390	0.209	1.120
3	644	0.866	0.475	0.129	0.827
4	652	0.741	0.469	0.176	0.965
5	620	0.619	0.494	0.147	0.866
6	643	0.575	0.414	0.188	1.132
7	613	0.480	0.414	0.206	1.136
8	653	0.389	0.438	0.158	1.033
9	618	0.299	0.440	0.159	0.987
10	596	0.206	0.347	0.148	0.927

Diffraction-collection protocols that invoke fast prescan procedures to assess whether a reflection is significant above background, and worthy of scanning more accurately at a slower speed, are not suitable for this purpose. Similar problems may be encountered when an area detector is employed to measure diffraction data.

Although E magnitudes can be scaled by at least a half-dozen different methods, procedures which allow for anisotropic thermal modeling, should it prove to be significant, are to be preferred (Levy, Thiessen & Brown, 1970; Sheriff & Hendrickson, 1987; Blessing & Langs, 1988). Bayesian processing of the data to obtain optimal estimates for F and $\sigma(F)$ is to be recommended (French & Wilson, 1978), although it is not advisable to tax the method by applying it to high-resolution data shells which have fewer than, say, 20% of the data observed at the $3\sigma(F)$ level. The temptation to measure as many significant strong reflections as possible to minimize the difficulty in solving a structure may often cause one to measure high-resolution data shells in which fewer than 10% of the data are observable. However, it may be risky to merge sparsely observed high-angle data with lower resolution shells, particularly with regard to accurately estimating the strongest and weakest E magnitudes. To analyze a set of scaled E values it is useful to sort the data into equal population shells as a function of increasing $\sin\theta/\lambda$, and compute the average $|E|^2$, the percentage of data observed at the 3σ level, the fraction of E values less than, say, 0.70 or greater than 1.50, and compare these with their expected values as is indicated in Table 1. The results may indicate how well the weakest and strongest reflections are estimated, particularly in the higher resolution shells.

Departures from the point-atom model

Data-reduction routines that provide reasonable F and $\sigma(F)$ estimates may be used to infer a minimum bound on the errors in deriving individual E magnitudes. The actual errors in deriving a set of E values will be larger than indicated by counting statistics, and for some data sets the goal of producing E values that correspond to a point-atom structure may be much more difficult to achieve than with others.

This point can be illustrated by examining data recorded both at room temperature and at 115 K for the monoclinic $P2_1$ structure of gramicidin A (Langs, Smith, Courseille, Précigoux & Hospital, 1991), a peptide structure that required the determination of more than 300 light-atom positions. The room-temperature structure was refined by constrained least squares to $R = 0.158$ for 2942 $2\sigma(F)$ data to 1.5 Å resolution, the low-temperature set to $R = 0.150$ for 6869 data to 1.2 Å resolution. By this measure it would appear that the accuracy of the amplitude data for either set was not significantly different in their agreement with the refined model. What is surprising, however, is that although the stronger E amplitudes which are common to both data sets are in general consistent within this ~15% variation for the majority of the data, there appears to be a significant reordering among the triples with regard to their A values ($A_{\mathbf{h},\mathbf{k}} = 2|E_{\mathbf{h}}E_{-\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|/N^{1/2}$). Table 2 lists a comparison of the cosine values of the phase invariants, $\Phi_{\mathbf{h},\mathbf{k}} = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{\mathbf{k}-\mathbf{h}}$ for those triples whose A value exceeds 1.50 for each set of data. The distribution of A values for both temperatures is approximately the same and allows one to project an equivalent A-weighted expected cosine

$$\varepsilon(\cos\Phi_{\mathbf{h},\mathbf{k}}) = \sum A_{\mathbf{h},\mathbf{k}}[I_1(A_{\mathbf{h},\mathbf{k}})/I_0(A_{\mathbf{h},\mathbf{k}})]/\sum A_{\mathbf{h},\mathbf{k}} \quad (1)$$

of ~0.66 for either of these two lists, which is equivalent to an average deviation of 49° from the zero-phase estimate assumed in the tangent formula. The true values for the average A-weighted cosine invariants

$$\langle\cos\Phi_{\mathbf{h},\mathbf{k}}\rangle_{\text{calc}} = \sum A_{\mathbf{h},\mathbf{k}}\cos\Phi_{\mathbf{h},\mathbf{k}}/\sum A_{\mathbf{h},\mathbf{k}} \quad (2)$$

are in reality, however, 0.17 and 0.46 respectively, for the room- and low-temperature data sets. Lowering the temperature of the crystal reduces these average phase invariant errors from 80 to 63°, to approach more closely the 49° value expected of a randomly distributed point-atom structure. Whereas the phases of the room-temperature structure are divergent under tangent-formula phase refinement, those of the isomorphous low-temperature structure tend to converge.

The lesson that one should draw from this comparison is that the degree to which a set of E values

Table 2. *Σ_2 triples for the monoclinic $P2_1$ structure of gramicidin A ($N = 332$ atoms) crystallized from methanol*

Invariants are coded as $\Phi = \varphi(N1) + \varphi(N2) + \varphi(N3) + T$; $\varphi(-N_i) = -\varphi(N_i)$.

A	$N1$	$N2$	$N3$	T	$\cos\Phi$
(*a*) Room temperature					
3.54	1	−99	−2	0	0.975
2.86	1	−21	−12	0	−0.123
2.60	3	−5	11	π	0.997
2.36	2	−50	−13	0	−0.728
2.25	1	−7	−215	0	0.472
2.12	2	−98	−12	0	0.530
2.06	6	11	−17	0	0.609
2.02	2	6	−272	π	−0.982
1.94	1	−394	−8	0	0.826
1.86	4	−12	−36	0	−0.949
1.84	3	22	−35	0	0.957
1.78	10	13	−17	π	0.612
1.74	1	−25	−263	π	1.000
1.73	1	−53	−162	π	0.663
1.71	4	−89	9	π	−0.097
1.70	2	38	−163	π	−0.957
1.70	6	22	−38	0	0.991
1.68	1	7	−637	0	−0.683
1.68	4	99	−9	0	0.912
1.68	7	20	−40	0	0.972
1.66	9	−44	−12	0	−0.981
1.66	1	−3	−853	π	−0.644
1.64	12	−18	−24	0	−0.993
1.62	15	−22	−18	π	0.650
1.61	7	−30	−41	0	−0.173
1.60	5	−6	−164	π	0.671
1.59	2	−65	−148	0	0.998
1.59	1	−101	−125	0	−0.492
1.59	3	62	−39	π	0.837
1.59	2	−17	−381	0	−0.501
1.58	2	66	−150	0	0.798
1.58	4	7	−205	0	0.965
1.58	2	−116	80	π	−0.950
1.57	1	−158	−95	0	−0.959
1.57	2	−301	−29	0	−0.207
1.55	3	−53	52	π	0.019
1.54	12	−42	−13	0	0.671
1.53	3	18	−128	π	0.138
1.53	4	−22	−82	0	0.943
1.53	1	−161	−102	0	0.894
1.53	2	−59	−220	0	−0.961
1.52	6	21	−74	π	0.645
1.51	3	15	−149	0	−0.748
1.51	7	24	−70	0	−0.693
		$\varepsilon(\cos) = 0.661$, $\langle\cos\rangle_{\text{calc}} = 0.175$!			
(*b*) Low temperature (115 K)					
3.26	1	−3	−32	0	1.000
2.47	2	−19	−32	0	0.999
2.41	2	−141	−4	0	0.940
2.37	1	5	−93	0	0.994
2.36	6	26	−9	0	1.000
2.34	4	−22	−32	0	0.908
2.25	6	44	−7	0	1.000
2.17	2	22	−69	π	0.766
2.13	1	32	−70	0	1.000
2.11	2	32	−67	0	0.993
2.10	4	19	−69	π	0.977
2.06	11	36	−14	0	−0.977
2.03	2	−26	−87	0	0.329
2.02	3	−141	10	0	0.809
2.02	8	11	−60	0	−0.719
2.02	13	26	−19	0	−0.145
1.97	5	−75	−11	0	0.999
1.97	5	51	−24	π	0.710
1.95	5	−28	46	π	0.988
1.91	2	−5	−310	π	−0.246
1.87	3	5	−317	0	0.953
1.87	1	−13	257	π	0.971
1.87	5	−13	−92	π	0.942
1.85	10	32	−41	0	0.760
1.84	7	−17	75	π	−0.889
1.84	4	−299	5	π	−0.023
1.82	14	53	−21	0	−0.942
1.82	4	40	−93	0	0.579
1.81	9	16	−82	0	−1.000
1.80	2	10	−306	0	1.000

Table 2 (*cont.*)

A	$N1$	$N2$	$N3$	T	$\cos\Phi$
1.80	1	141	41	0	0.997
1.79	7	8	−143	0	0.502
1.79	4	−69	67	π	0.927
1.79	11	39	−40	0	−0.631
1.78	1	214	27	0	0.794
1.77	6	71	34	0	1.000
1.76	4	53	93	π	−0.247
1.76	25	32	−31	0	0.849
1.75	2	17	−255	0	0.055
1.75	7	22	89	0	0.993
1.75	3	68	−89	0	0.243
1.74	5	15	−123	0	0.860
1.73	5	141	−14	0	0.825
1.71	4	444	−5	0	0.814
1.70	8	103	19	0	−0.806
1.70	2	−51	159	π	0.993
1.69	4	299	−14	0	0.981
1.69	8	−103	−22	π	0.073
1.68	3	25	260	0	0.178
1.67	17	−29	−58	π	0.167
1.66	12	−13	127	π	0.735
1.65	7	32	99	0	1.000
1.64	10	41	−69	0	−0.916
1.64	2	444	−14	0	0.933
1.63	5	147	21	0	0.577
1.63	8	−32	−98	0	−0.149
1.63	5	14	−209	0	0.029
1.63	32	−43	36	π	0.698
1.63	19	49	−45	0	−0.784
1.62	30	47	−35	0	1.000
1.61	2	9	−590	0	0.977
1.61	16	−17	91	π	0.999
1.60	13	32	−87	0	0.910
1.59	7	192	−15	0	0.964
1.57	5	−25	178	π	0.823
1.57	17	39	−68	0	−0.592
1.55	1	299	−53	0	0.945
1.55	4	32	−293	0	0.999
1.55	2	−13	−579	0	0.092
1.54	4	13	−500	π	0.994
1.54	19	67	−45	π	0.692
1.53	14	−78	−49	0	0.915
1.52	11	72	−62	0	0.132
1.52	8	170	−28	0	0.833
1.52	2	−147	93	π	−0.964
1.51	8	31	−166	0	0.200
1.51	5	−137	−49	0	0.962
1.51	16	51	−73	0	0.609
1.51	32	40	−62	0	0.686
1.50	1	25	−526	0	−0.165
1.50	14	−41	93	0	0.809

$\varepsilon(\cos) = 0.664$, $\langle\cos\rangle_{calc} = 0.461$.

diverges from the theoretical point-atom model is a real measure of the non-uniformity of temperature factors of the atoms of the structure. This effect is separate and distinct from how uniformly or independently the atoms are distributed throughout the unit cell. In conclusion, if probabilistic phase invariants are used to phase a macromolecular structure that has no heavy atoms or anomalous scatterers, one will either have to succeed in reducing the r.m.s. temperature-factor fluctuation to less than what was observed for the room-temperature gramicidin data, or conversely, rely on non-invariant methods such as entropy-based calculations in order to succeed.

Connectivity among the phases

The success of most direct-methods procedures is highly dependent on the order and multiplicity of interactions in which phases are determined through phase-invariant relationships that are probabilistically correct. Diffraction data that have been accurately measured, precisely scaled, and correspond reasonably well to a point-atom structure, should provide three-phase triples invariants for which the probabilistic estimates are quite reliable as a function of their A magnitudes. The tangent formula is expected to be able to provide a solution whose individual phase errors can be estimated from the α value computed from their total number of triples contributors at the end of the refinement

$$\alpha_{\mathbf{h}} = [(\textstyle\sum_k A_{\mathbf{h},\mathbf{k}}\cos\Phi_{\mathbf{h},\mathbf{k}})^2 + (\sum_k A_{\mathbf{h},\mathbf{k}}\sin\Phi_{\mathbf{h},\mathbf{k}})^2]^{1/2}$$

$$\alpha_{\mathbf{h}} \cong \textstyle\sum_k A_{\mathbf{h},\mathbf{k}} I_1(A_{\mathbf{h},\mathbf{k}})/I_0(A_{\mathbf{h},\mathbf{k}}) \qquad (3)$$

where the variance was shown to be (Karle & Karle, 1966)

$$\alpha^2(\varphi_{\mathbf{h}}) = \pi^2/3 + 1/I_0(\alpha)[\sum_{n=1}^{\infty} I_{2n}(\alpha)/n^2 - 4\sum_{n=0}^{\infty} I_{2n+1}(\alpha)/(2n+1)^2] \qquad (4)$$

and may be shown to produce r.m.s. phase values of $\sigma(\varphi_{\mathbf{h}}) \cong 73$, 50, 38, 31 and 21° for α values of 1.0, 2.0, 3.0, 4.0 and 8.0 respectively.

Generally speaking, if a larger structure is to be solved, the α values of its phases at the beginning of a convergence mapping must steadily increase as subsequent phases are determined, and build to a maximum of 5.0 or 6.0 or more at the midpoint of the map, before subsiding to values of less than, say 2.0, at which point additional phases in the tail of the map cannot be reliably determined.

In an iterative stepwise phase-extension process, if the error in pivotal, newly determined phases can be kept to less than ± 45°, that is if they are determined with an α value exceeding, say ~ 2.5, the procedure has a good chance of succeeding. But if early in this process, there are numerous weak linkages with α values less than 2.5 affecting pivotal phases, this process will be prone to failure. For most structures containing 100 or more non-hydrogen light atoms, iterative phase-extension processes will often not succeed unless an unusually large basis set of phases is employed to ensure against weak α linkages early in the phase-extension process. Global phase-refinement procedures (Yao, 1981), however, which use all the phase invariants from the beginning of the calculation, are not affected by these weak links.

To summarize this section, if one can produce a convergence mapping that builds to an α of at least 5.0, or preferably 6.0 or more for the majority of phases, and the data have been accurately measured and scaled and adequately model a point-atom structure, one is virtually assured of a convergence to an acceptably stable phase solution; that is, provided

one can randomly sample phase space thoroughly enough to find a starting point from which a solution can be found. However, if the convergence-map profile peaks out with maximum α values of 4.0 or less, traditional tangent-formula phasing procedures will most likely not be able to find a stable convergent solution. One should not be encouraged to expend a lot of time and resources to disprove this advice unless one intends to use some new technique that may be more robust than the traditional tangent-formula algorithms coded into most direct-methods programs. It remains to be shown to what extent a number of newer approaches that are being developed can converge to a solution under these circumstances.

Aberrant phase invariants

The major reason why tangent-formula methods cannot easily produce solutions for some structure determinations that have strong convergence maps (apart from unrecognized twinning and misidentified crystal symmetries) is the influence of aberrant pivotal phase invariants. This is clearly a serious problem for basis-set extension methods, and although random phasing procedures can phase around aberrant weak links in a convergence map, it may require a larger number of trials than normal to do so. Some techniques will be demonstrated in this section that may be useful for identifying certain of these troublesome invariants for large complex structures.

Algebraic formulae to evaluate the cosine values of the three-phase invariants were first derived 35 years ago (Karle & Hauptman, 1957; Vaughan, 1958), but applications were not immediate, both due to the computational expense at that time and the recognition that the accuracy of these calculations would rapidly decrease as more complex structures exhibited a greater number of interatomic vectors that overlapped by chance in the Patterson function. A number of efforts to improve these formulae to minimize the effect of Patterson overlap were undertaken (Hauptman, 1964; Hauptman, Fisher, Hancock & Norton, 1969; Fisher, Hancock & Hauptman, 1970; Karle, 1970; Busetta, 1977; Giacovazzo, 1977).

Two well known variants that were developed and actively employed include the TPROD (Hauptman, Fisher, Hancock & Norton, 1969)

$$|E_{\mathbf{h}}E_{-\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{\mathbf{k}-\mathbf{h}}) \simeq R_3 + K\Psi \quad (5)$$

$$R_3 = (1/4N^{1/2})[(3/2)(|E_{\mathbf{h}}E_{\mathbf{k}}|^2 + |E_{\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|^2 + |E_{\mathbf{h}}E_{\mathbf{k}-\mathbf{h}}|^2) + |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{\mathbf{k}-\mathbf{h}}|^2 - 7/2]$$

$$\Psi = \langle(|E_{\mathbf{l}}|^{1/2} - \xi)(|E_{\mathbf{l}-\mathbf{k}}|^{1/2} - \xi)(|E_{\mathbf{l}-\mathbf{h}}|^{1/2} - \xi)\rangle_{\mathbf{l}}$$

$$\xi = \langle|E_{\mathbf{h}}|^{1/2}\rangle_{\mathbf{h}}$$

and MDKS formulae (Fisher, Hancock & Hauptman, 1970)

$$|E_{\mathbf{h}}E_{-\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{\mathbf{k}-\mathbf{h}}) \simeq M(D - KS) \quad (6)$$

$$D = \langle(|E_{\mathbf{l}-\mathbf{k}}|^2 - 1)\,\big|\,|E_{\mathbf{l}}|,\ |E_{\mathbf{l}-\mathbf{h}}| \geq t\rangle_{\mathbf{l}}$$

$$S = \langle(|E_{\mathbf{l}-\mathbf{k}}|^2 - 1)\,\big|\,|E_{\mathbf{l}}| \geq t\rangle_{\mathbf{l}}\langle(|E_{\mathbf{l}-\mathbf{k}}|^2 - 1)\,\big|\,|E_{\mathbf{l}-\mathbf{h}}| \geq t\rangle_{\mathbf{l}},$$

where M and K are scaling constants to fit the distribution of calculated cosine values, and t is a conditional threshold value, say ~ 1.5, placed on certain of the magnitudes which must be satisfied before the associated terms can be included in the average. Although the TPROD and MDKS formulae are fairly well known, the only commercially available direct-methods program to incorporate these procedures has been *MITHRIL* (Gilmore, 1984) with the caveat that the estimates may on occasion prove to be unreliable (Gilmore, 1991).

A third less known variant called the ν-STAT formula (Langs, 1972)

$$\nu^+ = \frac{\text{No. of quadrupoles}(|E_{\mathbf{l}-\mathbf{k}}| \geq t_1\,\big|\,|E_{\mathbf{l}}|, |E_{\mathbf{l}-\mathbf{h}}| \geq t_1)}{\text{No. of quadrupoles}(|E_{\mathbf{l}-\mathbf{k}}|\text{obs}\,\big|\,|E_{\mathbf{l}}|, |E_{\mathbf{l}-\mathbf{h}}| \geq t_1)}$$

$$\nu^- = \frac{\text{No. of quadrupoles}(|E_{\mathbf{l}-\mathbf{k}}| \geq t_1\,\big|\,|E_{\mathbf{l}}| \leq t_2, |E_{\mathbf{l}-\mathbf{h}}| \geq t_1)}{\text{No. of quadrupoles}(|E_{\mathbf{l}-\mathbf{k}}|\text{obs}\,\big|\,|E_{\mathbf{l}}| \leq t_2, |E_{\mathbf{l}-\mathbf{h}}| \geq t_1)}$$

$$\nu(\text{random}) = (\text{No. of } |E|\text{'s} \geq t_1)/(\text{total No. of } |E|\text{'s}) \quad (7)$$

has been productively used on numerous difficult structures in the author's laboratory for nearly 20 years. Here an upper threshold t_1 is chosen to select E values greater than, say 1.75, for a positive cosine indicator (ν^+) and a lower threshold t_2 is set to select a similar number of the weakest E values and quadrupoles relationships for a negative cosine indicator (ν^-). The triples phase invariants are determined to be reliable ($\cos\Phi \simeq +1$) when the magnitudes of their computed frequencies have the order $\nu^+ > \nu(\text{random}) > \nu^-$, and may be suspected of being aberrant ($\cos\Phi \simeq -1$) when $\nu^+ < \nu(\text{random}) < \nu^-$. As with the TPROD and MDKS estimates, strong positive indications are usually correct for triples whose A values exceed 1.5, and less so as A values approach 1.0. Strong negative estimates are less certain for indicating aberrant triples; usually only about 25% of the triples flagged as probably being aberrant by the TPROD, MDKS or ν-STAT formulae will actually cause phasing problems. The simplest strategy to deal with this situation has been to exclude all of these triples from the convergence mapping, knowing that the majority of the aberrant invariants will be eliminated at the cost of losing a small fraction of the reliable triples available in the data set.

Table 3. *Σ_1 triples and zonal restricted Σ_2 triples for the orthorhombic $P2_12_12_1$ structure of gramicidin A crystallized from ethanol (N = 334 atoms)*

Aberrant triples are indicated by an asterisk at the extreme right-hand side of the table. The calculation used 1102 E's $\geq t_1 = 1.75$, 1500 E's $\leq t_2 = 0.32$, ν(random) = 0.051.

h			k			k − h			$N1$	$N2$	$N3$	ν^+	ν^-	A	T
(*a*) Σ_1 triples															
0	27	2	0	−27	2	0	0	−4	1	−1	−348	0.066	0.039	1.96	π
2	0	1	−2	0	1	0	0	−2	3	−3	−9	0.096	0.063	1.91	0
16	0	1	−16	0	1	0	0	−2	4	−4	−9	0.054	0.047	1.83	0
2	27	3	2	−27	3	−4	0	−6	8	−8	−2	0.032	0.043	1.69	0
3	6	2	3	−6	2	−6	0	−4	5	−5	−19	0.033	0.038	1.41	0*
0	33	1	0	−33	1	0	0	−2	12	−12	−9	0.059	0.033	1.18	0
0	25	1	0	−25	1	0	0	−2	15	−15	−9	0.057	0.055	1.14	0
8	0	1	−8	0	1	0	0	−2	36	−36	−9	0.054	0.040	0.77	0
17	0	1	−17	0	1	0	0	−2	39	−39	−9	0.024	0.042	0.76	π*
0	0	2	0	0	2	0	0	−4	9	−9	−348	0.056	0.078	0.74	0*
2	0	5	−2	0	5	0	0	−10	16	−16	−122	0.052	0.042	0.72	0
2	35	3	2	−35	3	−4	0	−6	77	−77	−2	0.067	0.031	0.69	0
0	35	2	0	−35	2	0	0	−4	11	−11	−348	0.036	0.038	0.69	π
2	14	3	2	−14	3	−4	0	−6	108	−108	−2	0.014	0.034	0.61	π*
0	8	2	0	−8	2	0	0	−4	21	−21	−348	0.033	0.049	0.57	0*
0	8	2	0	8	−2	0	−16	0	21	−21	−350	0.033	0.038	0.57	0
15	0	2	−15	0	2	0	0	−4	24	−24	−348	0.033	0.048	0.56	π
3	14	2	3	−14	2	−6	0	−4	62	−62	−19	0.020	0.033	0.55	0*
5	0	5	−5	0	5	0	0	−10	34	−34	−122	0.031	0.040	0.52	π*
(*b*) Zonal restricted Σ_2 triples															
4	0	6	−2	0	−1	−2	0	−5	2	−3	−16	0.062	0.047	3.78	0
0	27	2	0	0	2	0	−27	−4	1	9	−35	0.055	0.047	3.53	0
0	27	2	0	8	2	0	−35	−4	1	21	−40	0.063	0.031	3.05	0
4	0	6	16	0	−1	−20	0	−5	2	4	−72	0.036	0.038	2.79	0
0	27	2	0	−35	−2	0	8	0	1	−11	161	0.055	0.028	2.70	0
2	0	1	16	0	−1	−18	0	0	3	4	−88	0.028	0.040	2.69	0
4	0	6	−2	0	1	−2	0	−7	2	−3	178	0.056	0.044	2.51	0
2	0	1	−6	0	−4	4	0	3	3	−19	32	0.035	0.052	2.47	0
2	0	1	15	0	−2	−17	0	1	3	24	39	0.038	0.042	2.32	π
0	33	1	0	−25	1	0	−8	−2	12	−15	−21	0.030	0.037	2.24	0
0	27	2	0	−26	13	0	−1	−15	1	−27	−366	0.035	0.041	2.19	π
0	27	2	0	−13	−6	0	−14	4	1	−48	−124	0.029	0.031	2.18	0
0	0	2	0	35	2	0	−35	−4	9	11	−40	0.043	0.042	2.09	0
16	0	1	0	−25	−1	−16	25	0	4	−15	93	0.029	0.041	2.08	π
2	0	1	0	−33	−1	−2	33	0	3	−12	135	0.049	0.026	2.06	π
0	27	2	0	−33	−1	0	6	−1	1	−12	1045	0.048	0.038	2.03	π
0	27	2	0	−25	−1	0	−2	−1	1	−15	−1028	0.035	0.032	2.01	0
4	0	6	−6	0	−4	2	0	−2	2	−19	−110	0.036	0.050	2.00	0
2	0	1	0	−25	−1	−2	25	0	3	−15	159	0.043	0.042	1.97	π
0	35	2	0	−8	2	0	−27	−4	11	−21	−35	0.041	0.034	1.91	0
18	0	12	−15	0	2	−3	0	−14	7	−24	−52	0.030	0.037	1.91	π
2	0	1	4	0	−3	−6	0	2	3	32	84	0.035	0.045	1.90	0
4	0	6	0	0	2	−4	0	−8	2	9	−442	0.053	0.052	1.85	0*
2	0	5	6	0	−4	−8	0	−1	16	19	−36	0.037	0.038	1.85	0*
0	27	2	0	−26	−13	0	−1	11	1	−27	−1055	0.065	0.035	1.84	0
0	27	2	0	−27	4	0	0	−6	1	−35	−620	0.051	0.027	1.83	π
3	6	0	−1	27	0	−2	−33	0	6	14	−135	0.037	0.029	1.83	π
2	0	1	−8	0	1	6	0	−2	3	−36	−84	0.054	0.040	1.83	0
2	0	1	−4	0	−3	2	0	2	3	−32	110	0.043	0.054	1.81	0*
2	0	1	15	0	2	−17	0	−3	3	24	−210	0.035	0.046	1.81	0
2	0	1	−2	0	−5	0	0	4	3	−16	348	0.067	0.052	1.78	0
2	0	1	−5	0	9	3	0	−10	3	−31	−139	0.038	0.037	1.77	0
4	0	6	−5	0	9	1	0	15	2	−31	−149	0.044	0.045	1.76	π
0	0	2	−6	0	−4	6	0	2	9	−19	84	0.038	0.047	1.73	0
0	5	4	0	8	2	0	−13	−6	17	21	−48	0.017	0.032	1.73	0*
1	27	0	5	6	0	−6	−33	0	14	22	−54	0.038	0.036	1.71	0
16	0	1	−5	0	9	−11	0	−10	4	−31	−170	0.035	0.037	1.69	π
18	0	12	−15	0	−2	−3	0	−10	7	−24	−139	0.035	0.018	1.64	0
0	27	2	0	0	−10	0	−27	8	1	−122	−379	0.046	0.039	1.63	π
2	0	1	−2	0	5	0	0	−6	3	−16	−620	0.073	0.040	1.63	0
0	0	2	2	0	5	−2	0	−7	9	−16	−178	0.056	0.044	1.62	0
2	0	5	−5	0	9	3	0	−14	16	−31	−52	0.054	0.026	1.60	0
2	0	1	−2	0	9	0	0	−10	3	−56	−122	0.062	0.042	1.59	0
2	0	1	−3	0	14	1	0	−15	3	−52	−149	0.030	0.040	1.58	π
0	27	2	0	−13	−5	0	−14	3	1	−68	−805	0.028	0.032	1.58	π
0	33	1	0	−25	−1	0	−8	0	12	−15	−161	0.045	0.036	1.57	0
0	0	2	0	−8	−2	0	8	0	9	−21	161	0.038	0.042	1.55	0
0	27	2	0	−27	−6	0	0	4	1	−216	348	0.048	0.040	1.54	0
2	0	1	3	0	14	−5	0	−15	3	52	−191	0.040	0.058	1.54	0
4	0	6	2	0	5	−6	0	−11	2	16	−998	0.048	0.038	1.51	0*
4	0	6	−6	0	−8	2	0	2	2	−95	110	0.028	0.040	1.50	0*
16	0	1	2	0	−5	−18	0	4	4	16	920	0.044	0.029	1.49	0
2	0	5	4	0	−3	−6	0	−2	16	32	−84	0.034	0.042	1.47	0*
0	0	2	0	13	6	0	−13	−8	9	48	−83	0.032	0.042	1.46	0
4	0	6	−15	0	−2	11	0	−4	2	−24	−894	0.023	0.039	1.45	π*
0	27	2	0	−26	−17	0	−1	15	1	−308	−366	0.051	0.032	1.45	0
2	0	1	−7	0	−2	5	0	1	3	−37	448	0.042	0.043	1.45	0*
16	0	1	2	0	−9	−18	0	8	4	56	227	0.013	0.034	1.44	0

Table 3 (*cont.*)

h			k			k − h			N1	N2	N3	ν'	ν	A	T
2	0	5	4	0	3	−6	0	−8	16	32	−95	0.040	0.037	1.44	0
16	0	1	−6	0	4	−10	0	−5	4	−19	927	0.024	0.030	1.44	0
18	0	12	−7	0	−2	−11	0	−10	7	−37	−170	0.024	0.032	1.42	0
6	0	4	−15	0	2	9	0	−6	19	−24	−136	0.012	0.040	1.42	0
2	0	5	−4	0	−3	2	0	−2	16	−32	−110	0.054	0.039	1.41	0
2	0	5	15	0	−2	−17	0	−3	16	24	−210	0.022	0.027	1.40	π
4	0	6	6	0	4	−10	0	−10	2	19	−1262	0.034	0.043	1.39	0*
0	0	2	0	27	4	0	−27	−6	9	35	−216	0.042	0.052	1.35	0*
2	0	1	2	0	−9	−4	0	8	3	56	442	0.060	0.039	1.34	0
0	0	2	17	0	1	−17	0	−3	9	39	−210	0.034	0.037	1.34	0
2	0	5	−5	0	5	3	0	−10	16	−34	−139	0.043	0.044	1.33	0
18	0	12	2	0	5	−20	0	−17	7	16	−911	0.037	0.031	1.32	0
0	27	2	0	−27	−8	0	0	6	1	−379	620	0.049	0.028	1.30	0
16	0	1	−2	0	9	−14	0	−10	4	−56	−493	0.033	0.041	1.30	0
4	0	6	7	0	−2	−11	0	−4	2	37	−894	0.022	0.047	1.29	π*
4	0	6	11	0	10	−15	0	−16	2	170	−194	0.017	0.026	1.29	0
16	0	1	−5	0	−5	−11	0	4	4	−34	894	0.027	0.035	1.29	π
16	0	1	−1	0	15	−15	0	−16	4	−149	−194	0.037	0.028	1.28	π
6	0	4	−4	0	3	−2	0	−7	19	−32	−178	0.028	0.036	1.27	0*
4	0	6	−3	0	14	−1	0	−20	2	−52	−707	0.026	0.039	1.27	π
18	0	12	−20	0	−5	2	0	−7	7	−72	−178	0.009	0.038	1.27	π*
0	0	2	−2	0	−9	2	0	7	9	−56	178	0.044	0.044	1.26	0
2	0	5	−20	0	−5	18	0	0	16	−72	88	0.025	0.034	1.25	0

An example of the application of the ν-STAT cosine analysis for the phase-restricted triples produced by the orthorhombic $P2_12_12_1$ structure of gramicidin A (Langs, 1988) is presented in Table 3. The cosine estimates are remarkably good considering this structure contains more than 300 atoms in the asymmetric unit and its Patterson map exhibits a horrendous amount of peak overlap resulting from the helical conformation of the molecular dimer. Note that among the 100 triples listed in Table 3, 21 of these triples (7 of 19 Σ_1 and 14 of 81 Σ_2 triples) are shown to be aberrant, *i.e.* $\cos\Phi = -1.0$. Although the identity of these 21 triples was not known *a priori*, inspection of the list clearly indicates that only two of these 21 triples have values of ν^+ exceeding ν^-.

In the actual solution of this structure a significant portion of the starting basis set of 42 phases was selected by assuming as correct, all those triples indications for which $\Delta\nu = \nu^+ - \nu^-$ exceeded 0.005. A total of 33 zonal restricted phases could be defined and linked in terms of only five symbols in this manner as indicated in Table 4. In this starting set of 33 + 9 other phases, only phase number 998 proved to be in error when the correct values of the permuted symbols were tested. A solution was found for which $\simeq 175$ atoms of the structure could be elaborated by *RANTAN* fragment recycling (Yao, 1983); difference Fourier maps revealed the remaining atoms of the structure which summed to a total of 334 non-hydrogen, full and partial occupancy sites. Although the account of this particular structure determination seems highly straightforward and trivial, one should be assured that these remarkable results could not have been obtained if the data had been carelessly measured or poorly scaled.

Another unexpected advantage of the ν-STAT formula is that the ν^+ and ν^- estimators normally

Table 4. *Basis set of* 33 *zonal restricted phases for gramicidin A as determined from Table* 3

Five symbols (a, b, c, d, $\alpha/2$) and eight symbolic relationships ($\varphi_i = \varphi_j \pm \varphi_k$ coded as $i = j \pm k$) are employed to express this set.

No.	φ			No.	φ	No.	φ	No.	φ
1	π/2	(0,27,2)	Origin	16	−π/2	620	π	216	−π/2
3	π/2	(2,0,1)	and	35	π/2	14	π/2	998	−π/2*
6	π/2	(3,6,0)	enantiomer	21	a	379	−π/2	4	α/2
12	π/2	(0,33,1)		40	= 21 + 1	84	b	920	= 4 + 16
				11	= 21 + 35	36	= 3 − 84	442	π
2	+	Σ_1		161	= 11 − 1	52	c	27	d
9	+	phases		178	−π/2	31	= 16 − 52	1055	= 1 − 27
122	+			135	+	56	π/2		
348	π			1045	= π	15	= 12 − 161		

* Reflection number 998 is in error and has a true phase value of 90°.

do not have to be markedly rescaled to fit the expected cosine distribution. An important observation has been that on those rare occasions that a set of data has produced a lopsidedly large percentage of aberrant ν-STAT triples estimates, the E data have either been shown to be badly scaled, thermally unrepresentative of a point-atom model, incorrectly indexed on the wrong cell, or the crystal gave evidence of being twinned. Thus if one has had a problem in solving a particular crystal structure by direct phasing methods, it is highly worthwhile to perform a ν-STAT triples analysis if for no other reason than to confirm that certain of these effects may be plaguing your data and efforts should be taken to obtain a better set.

Aberrance owing to translational symmetry

Some of the phasing traps that occur in a triples listing are a consequence of translational symmetry associated with screw axes and glide planes. A simple example of this symmetry-required aberrance could be the conflict between two Σ_1 indications, 0 and π, for the same phase. One indication must be correct,

Table 5. *Example of a $P3_2$ structure that possesses inconsistent triples,* [D-*Hyi*^{2},L-*Hyi*^{4}]-*meso-valinomycin* (*N* = 92 *atoms*)

There are no phase-restricted triples in this space group. Most of the pairs of triples shown below are either consistent or inconsistent, depending on the value of T (°).

h			−k			k − h			N1	N2	N3	ν'	ν	A	T (°)	cosΦ
1	3	2	−3	−13	1	2	10	−3	1	11	10	0.037	0.032	3.31	120	0.84*
1	3	2	2	−13	−3	−3	10	1	1	11	10	0.030	0.054	3.31	240	−0.88*
2	3	2	−4	−3	2	2	0	−4	3	125	4	0.072	0.039	2.24	0	0.97
2	3	2	2	−3	−4	−4	0	2	3	125	4	0.077	0.032	2.24	0	0.97
2	0	2	2	−1	−2	−4	1	0	4	20	133	0.080	0.042	1.91	240	0.98†
2	0	2	−2	1	2	0	−1	−4	4	−20	−133	0.087	0.050	1.91	240	0.87†
1	3	2	2	−2	−3	−3	−1	1	1	134	107	0.044	0.057	1.71	120	−0.49
1	3	2	−3	−2	1	2	−1	−3	1	134	107	0.028	0.055	1.71	240	−0.51
1	−13	2	2	−11	−3	−3	24	1	11	30	173	0.043	0.031	1.58	240	0.89
1	−13	2	−3	−11	1	2	24	−3	11	30	173	0.038	0.056	1.58	120	0.05
0	−20	6	−3	20	0	3	0	−6	28	48	155	0.054	0.050	1.41	240	0.43†
0	−20	6	3	20	−3	−3	0	−3	28	48	−155	0.055	0.046	1.41	120	0.74†
1	10	2	2	−11	−3	−3	1	1	10	30	291	0.044	0.040	1.40	0	−0.98
1	10	2	−3	−11	1	2	1	−3	10	30	291	0.043	0.044	1.40	0	−0.98
1	0	1	1	−23	−2	−2	23	1	31	70	158	0.061	0.051	1.29	120	0.01
1	0	1	−2	−23	1	1	23	−2	31	70	158	0.098	0.043	1.29	240	0.86
0	−1	2	1	1	−1	−1	0	−1	20	356	−31	0.100	0.044	1.28	240	0.98†
0	−1	2	−1	1	0	1	0	−2	20	356	31	0.103	0.040	1.28	120	0.97†

* These two triples were the strongest in the entire $P3_2$ data set yet it is physically impossible for them both to be correct. The value of the first invariant exactly equals that of the second in the list minus 120°.

† Note that these particular pairs of triples are not symmetry-related and their values are totally independent of one another. The calculation used 212 E's $\geq t_1 = 1.75$, 300 E's $\leq t_2 = 0.25$, ν(random) = 0.061.

the other must be wrong, and one can never be absolutely certain which is which unless one solves the structure.

Inconsistent/consistent triples

Certain trigonal, tetragonal, hexagonal and cubic space groups exhibit potentially inconsistent triples, that is two triples which involve the same three parent reflections, but in a non-identical symmetry-related manner (Han & Langs, 1988). The structure of [D-Hyi^2,L-Hyi^4]-*meso*-valinomycin ($P3_2$, $N = 92$ atoms, Pletnev, Mikhailova, Ivanov, Langs, Grochulski & Duax, 1991) is chosen to clarify further the nature of these special kinds of triples. This structure is quite unusual in that the two strongest triples, those with the largest A values, actually represent an inconsistent pair, as is indicated in Table 5. The phase invariant for the first triple is $\Phi_1 = \varphi_1 + \varphi_{10} + \varphi_{11} + 120°$, while that of the second is $\Phi_2 = \varphi_1 + \varphi_{10} + \varphi_{11} + 240°$; *i.e.* the two phase indications are inconsistent since they must disagree by 120°. This structure cannot be easily solved if one ignores this fact, since the average of the two invariants will impart a 92° phase error through the tangent formula. The ν-STAT estimators computed for these two triples are not the same, which demonstrates that the two invariant estimates are not identical, and that the second triple is more likely aberrant. Removing the second triple from the list reduces the tangent-formula error from 92 to only 32°. Please note in Table 5, that in addition to three other inconsistent pairs of triples (7 and 8, 9 and 10, 15 and 16), there are two consistent pairs as indicated by triples 3 and 4 and 13 and 14, and other curious pairs of triples, indicated by †, which are neither consistent nor inconsistent because of a change of sign associated with one of the phases. One should readily conclude that the ν-STAT indications given in the table can be used to advantage to avoid the phasing traps posed by the symmetry-induced aberrance of this space group.

Inconsistent/consistent quadrupoles

Aberrant phase indications are further propagated through quadrupole relationships among the triples. Normal quadrupole relationships are formed when the invariant values of four triples sum to zero as a consequence of phases being paired with their Friedel mates.

$$\Phi_1 = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{\mathbf{k}-\mathbf{h}}$$
$$\Phi_2 = \varphi_{\mathbf{k}} - \varphi_{\mathbf{l}} + \varphi_{\mathbf{l}-\mathbf{k}}$$
$$\Phi_3 = \varphi_{\mathbf{l}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}-\mathbf{l}}$$
$$\Phi_4 = -\varphi_{\mathbf{k}-\mathbf{h}} - \varphi_{\mathbf{l}-\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{l}}$$
$$\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 = 0 \pmod{2\pi}. \qquad (8)$$

When four strong triples enter into a normal quadrupole it is usually the case that each invariant $\Phi_i \cong 0$. It should also be clear that when an aberrant invariant, $\Phi_i \cong \pi$, enters into such a quadrupole, the other three invariants must sum together to equal this departure from zero. In the case of centrosymmetric structures aberrant triples must occur in pairs, such that if one of the triples is known to be π, either one or possibly all three of the remaining triples must also be aberrant π invariants.

Table 6. *Examples of a structure that exhibits troublesome π quadrupoles, ternatin $P2_12_12_1$ ($N = 106$ atoms)*

404 E's $\geq t_1 = 1.75$, 499 E's $\leq t_2 = 0.26$, ν(random) = 0.064. The ν-STAT values clearly indicate one aberrant triple in the first π quadrupole, and more remarkably, three aberrant triples in the second and third quadrupoles. It may become increasingly necessary to employ this aberrant phase information in more complex structure determinations.

	h			**−k**			**k − h**			$N1$	$N2$	$N3$	ν^+	ν^-	A	T
4	0	3	−1	0	8	−3	0	−11	6	−8	−111	0.009	0.055	2.78	π^*
4	0	3	1	0	8	−5	0	−11	6	8	−132	0.051	0.037	2.73	0
1	0	19	−4	0	−8	3	0	−11	5	−9	−111	0.020	0.021	2.67	0
1	0	19	4	0	−8	−5	0	−11	5	9	−132	0.042	0.048	2.62	0
6	0	6	4	0	1	−10	0	−7	7	29	−123	0.023	0.063	2.05	0*
7	0	6	4	0	1	−11	0	−7	1	29	−465	0.048	0.051	2.04	0
6	0	6	4	0	−13	−10	0	7	7	113	123	0.009	0.056	1.69	π^*
7	0	6	4	0	−13	−11	0	7	1	113	465	0.028	0.038	1.68	0*
4	0	3	−1	0	8	−3	0	−11	6	−8	−111	0.009	0.055	2.78	π^*
1	0	19	−4	0	3	3	0	−22	5	6	427	0.044	0.040	2.27	0
1	0	8	−3	0	−11	2	0	3	8	−111	248	0.026	0.049	1.44	0*
1	0	19	2	0	3	−3	0	−22	5	248	−427	0.027	0.069	1.18	0*

Translational symmetry also allows for consistent and inconsistent quadrupoles (Viterbo & Woolfson, 1973), for example in the structure of the cyclic *N*-methyl amino heptapeptide ternatin ($P2_12_12_1$, $N =$ 106) as indicated in Table 6. In this particular space group there are certain additional non-Friedel-related quadrupole relationships for which the sum of the four triples invariants must either equal 0 (consistent) or π (inconsistent). The ν-STAT estimators are given in Table 6 for three π quadrupoles exhibited by this structure. The ν^+, ν^- estimators suggest that the first triple in the first π quadrupole, and the first, third and fourth triples in the second and third π quadruples are aberrant π invariants. Note also, that there are two distinct arrangements of phases in the π quadruples indicated. Each triple in a normal quadrupole (8) is comprised of three phases, and each of those three phases appears as one of the phases in each of the three remaining triples. That is for Φ_1, $\varphi_{\mathbf{k}}$ appears in Φ_2, $\varphi_{\mathbf{h}}$ in Φ_3, and $\varphi_{\mathbf{h}-\mathbf{k}}$ in Φ_4. This phase pattern occurs for only the third π quadrupole in Table 6, but the first two π quadrupoles exhibit a different pattern than previously had been noted (Langs & Han, 1988).

Prognosis

As with all scientific endeavours, there must be practical and experimental limitations on what can be achieved with any developing technology. Direct phasing methods are undergoing changes that may markedly increase the level of complexity of structures which may be determined by these procedures. The object of this presentation was to call attention to some important considerations that are often ignored in a computational assault on a structure determination. Indeed, there is a danger that some of the wars that will be waged will be futile if the will of the data is not heard. Hopefully other suggestions offered in the course of this symposium will also temper our judgment as we venture forth. Yet one cannot be totally prepared for the unknown, and macromolecular applications will undoubtably test our will in many other unexpected ways.

This research has been supported in part by NIH grants GM32812 and GM46733.

References

BLESSING, R. H. & LANGS, D. A. (1988). *Acta Cryst.* A**44**, 729–735.

BUSETTA, B. (1977). *Acta Cryst.* A**33**, 441–444.

COCHRAN, W. (1958). *Acta Cryst.* **11**, 579–585.

COCHRAN, W. & WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 450–451.

FISHER, J., HANCOCK, J. & HAUPTMAN, H. (1970). Report 7157. Naval Research Laboratory, Washington, DC, USA.

FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* A**34**, 517–525.

GIACOVAZZO, C. (1977). *Acta Cryst.* A**33**, 527–531.

GILLIS, J. (1948). *Acta Cryst.* **1**, 174–179.

GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42–46.

GILMORE, C. J. (1991). In *Direct Methods of Solving Crystal Structures*, edited by H. SCHENK, pp. 177–186. New York: Plenum.

GILMORE, C. J., HENDERSON, A. N. & BRICOGNE, G. (1991). *Acta Cryst.* A**47**, 842–846.

GUO, D. Y. & HAUPTMAN, H. (1990). Proc. Am. Crystallogr. Assoc. Meet., New Orleans, LA, Abstract T02.

HAN, F. & LANGS, D. A. (1988). *Acta Cryst.* A**44**, 563–566.

HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70–75.

HARRISON, R. W. (1989). *Acta Cryst.* A**45**, 4–10.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421–1433.

HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* B**25**, 811–814.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*, ACA Monograph No. 3. Wilmington: The Letter Shop.

HAUPTMAN, H. & KARLE, J. (1954). *Acta Cryst.* **7**, 369–374.

KARLE, I. L., HAUPTMAN, H., KARLE, J. & WING, A. B. (1958). *Acta Cryst.* **11**, 257–263.

KARLE, I. L. & KARLE, J. (1964). *Acta Cryst.* **17**, 835–841.

KARLE, I. L. & KARLE, J. (1966). *Acta Cryst.* **21**, 860–868.

KARLE, I. L., KARLE, J. & ESTLIN, J. A. (1967). *Acta Cryst.* **23**, 494–500.

KARLE, I. L., KARLE, J., MASTROPAOLO, D., CAMERMAN, A. & CAMERMAN, N. (1983). *Acta Cryst.* B**39**, 625–637.

KARLE, J. (1970). *Acta Cryst.* B**26**, 1614–1617.

KARLE, J. (1971). *Acta Cryst.* B**27**, 2063–2065.

KARLE, J. (1989). *Acta Cryst.* A**45**, 765–781.

KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515–524.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.
KASPER, J. S., LUCHT, C. M. & HARKER, D. (1949). *Acta Cryst.* **3**, 436–455.
LANGS, D. A. (1972). Unpublished results, manuscript in preparation.
LANGS, D. A. (1988). *Science*, **241**, 188–191.
LANGS, D. A. & HAN, F. (1988). *Acta Cryst.* A**44**, 657–661.
LANGS, D. A., SMITH, G. D., COURSEILLE, C., PRÉCIGOUX, G. & HOSPITAL, M. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 5345–5349.
LEVY, H. A., THIESSEN, W. F. & BROWN, G. M. (1970). Proc. Am. Crystallogr. Assoc. Meet., New Orleans, LA, Abstract B6.
OH, Y. L. & MASLEN, E. N. (1968). *Acta Cryst.* B**24**, 883–897.
PLETNEV, V. Z., MIKHAILOVA, I. YU., IVANOV, V. T., LANGS, D. A., GROCHULSKI, P. & DUAX, W. L. (1991). *Biopolymers*, **31**, 409–415.
ROGERS, D. & WILSON, A. J. C. (1955). *Acta Cryst.* **8**, 365–366.
SHERIFF, S. & HENDRICKSON, W. A. (1987). *Acta Cryst.* A**43**, 118–121.
SJÖLIN, L. & PRINCE, E. (1991). Proc. Am. Crystallogr. Assoc. Meet., Toledo, OH, Abstract PA05.
SJÖLIN, L., PRINCE, E., SVENSSON, L. A. & GILLILAND, G. L. (1991). *Acta Cryst.* A**47**, 216–223.
TSOUCARIS, G. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 287–360. New York: Plenum.
VAND, V. & PEPINSKY, R. (1954). *Acta Cryst.* **7**, 451–452.
VAUGHAN, P. A. (1958). *Acta Cryst.* **11**, 111–115.
VITERBO, D. & WOOLFSON, M. M. (1973). *Acta Cryst.* A**29**, 205–208.
WOOLFSON, M. M. & YAO, J. (1990). *Acta Cryst.* A**46**, 409–413.
YAO, J. (1981). *Acta Cryst.* A**37**, 642–644.
YAO, J. (1983). *Acta Cryst.* A**39**, 35–37.